Highly reinforced structure of a C-terminal dimerization domain in von Willebrand factor

Yan-Feng Zhou and Timothy A. Springer

Program in Cellular and Molecular Medicine, Children's Hospital Boston and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA

Key Points

- C-terminal cystine knot monomers in VWF are highly elongated and form antiparallel dimers.
- Three disulfides across the dimer interface flanked by the cystine knots in each monomer form a highly force-resistant structure.

Introduction

Von Willebrand factor (VWF) is a mosaic of domains with many types of binding sites that crosslinks platelets to one another and to the vessel wall in hemostasis and thrombosis. Long length is key to VWF’s function as a sensor for flow changes at sites of hemostasis. Polymerization is mediated by specialized N- and C-terminal domains. Dimerization through a C-terminal cystine knot (CK) (CTCK) domain occurs first in the endoplasmic reticulum and is followed later in biosynthesis by N-terminal linkage through the D3 domain as VWF dimers assemble into helical tubules in the Golgi and Weibel-Palade bodies.

Domains with CKs contain a unique motif of 3 disulfide bonds. Two closely spaced cysteines on each of 2 polypeptide segments disulfide link to form a ring through which a third disulfide links other polypeptide segments crosses. The CK is actually a motif and not a domain; CKs are found in domains including knottins and cytokines that are otherwise completely unrelated.

CTCK domains were called C-terminal or CT domains by Bork, who found them at the C termini of the CCN (cysteine-rich, connective tissue growth factor, nephroblastoma overexpressed) family of regulatory matricellular proteins and the Slit family of proteins that regulate axon guidance. Bork recognized lesser sequence homology to domains at the C termini of VWF, some gel-forming mucins, hemolectin, and norrie disease protein (norrin). Eight of 10 or 11 cysteines were invariant in CT domains. Finally, Bork found still weaker homology, but conservation of the spacing of 6 of the cysteines, in CK cytokines such as transforming growth factor β (TGFβ).

Twenty years has passed since the CTCK domain family was identified and suggested to be structurally related to TGFβ. However, despite presence in a diverse group of functionally important proteins mutated in disease, the CTCK domain has been recalcitrant to structural elucidation. Early work on VWF showed that at least one inter-chain disulfide was present in the C-terminal region. Later chemical and mutational studies characterized the 11 cysteines present in the VWF CTCK domain. Intra-chain connections were defined for 8 cysteines, and among the 3 remaining cysteines, either 1 or 3 were proposed to mediate inter-chain dimerization. Here, we report the crystal structure of the VWF CTCK dimer. Our structure defines the principles for assembly of proteins with functions as diverse as hemostasis (VWF), barrier function (mucins), axonal guidance (Slit), and regulation of growth, migration, and differentiation (the CCN family and Norrie disease protein).

The C-terminal cystine knot (CK) (CTCK) domain in von Willebrand factor (VWF) mediates dimerization of proVWF in the endoplasmic reticulum and is essential for long multimers required for hemostatic function. The CTCK dimer crystal structure reveals highly elongated monomers with 2 β-ribbons and 4 intra-chain disulfides, including 3 in the CK. Dimerization buries an extensive interface of 1500 Å² corresponding to 32% of the surface of each monomer and forms a super β-sheet and 3 inter-chain disulfides. The shape, dimensions, and N-terminal connections of the crystal structure agree perfectly with previous electron microscopic images of VWF dimeric bouquets with the CTCK dimer forming a down-curved base. The dimer interface is suited to resist hydrodynamic force and disulfide reduction. CKs in each monomer flank the 3 inter-chain disulfides, and their presence in β-structures with dense backbone hydrogen bonds creates a rigid, highly crosslinked interface. The structure reveals the basis for von Willebrand disease phenotypes and the fold and disulfide linkages for CTCK domains in diverse protein families involved in barrier function, eye and inner ear development, insect coagulation and innate immunity, axon guidance, and signaling in extracellular matrices. (Blood. 2014;123(12):1785-1793)
Methods

Constructs and cell culture

Human VWF CTCK (residues 2720-2813 of pre-proVWF) was expressed in vector ETS. VWF sequence was fused C-terminal to the N-terminal signal peptide, His6 tag, and 3C protease site (LEVLFQGP). HEK293S N-acetylglucosaminyl transferase I-deficient cells were transfected using polyethyleneimine. Single colonies were selected in Dulbecco’s modified Eagle’s medium and 5% fetal bovine serum with 1 mg/mL G418. This study was conducted in accordance with the Declaration of Helsinki.

Characterization of the CTCK domain

CTCK was secreted as a disulfide-linked dimer (supplemental Figure 1, available on the Blood Web site). Secretion was greatly decreased if the His6 tag and 3C protease site were not included in the expression construct. Expression in supramatant was semiquantitatively measured by western blotting with anti-VWF polyclonal antibody (Delta Bio labs, Gilroy, CA). One N-glycosylation site in CTCK sequence was confirmed by endoglycosidase H (Roche) digestion (supplemental Figure 1).

Purification and crystallization

CTCK in square bottle suspension culture Ex-CELL 293 serum-free medium supernatant was purified using Ni-NTA (Qiagen) and Hitrap Q (GE Healthcare) chromatography as described for other C-terminal fragments of VWF, including buffers, pH, and salt gradients, except that Hitrap Q was substituted for mono Q. The ion exchange eluate was concentrated and exchanged into 20 mM Bicine, pH 6.2, 0.15 M NaCl, and 3 mM EDTA and digested with a mass ratio of 1:400 of endoglycosidase H: CTCK for 48 hours at 22°C. The digest was concentrated and loaded onto a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with 20 mM sodium acetate, pH 5.5, 0.15 M NaCl. Fractions containing CTCK dimer (supplemental Figure 1B) were concentrated to 20 mg/mL. Material was crystallized with the N-terminal His6 tag intact; material with the tag removed did not yield crystals.

Single crystals were obtained with a 1:1 volume to volume ratio well solution of 0.1 M bicine, pH 8.2 to 8.3, and 1.3 to 1.4 M ZnSO4 at 20°C in hanging drops. Crystalization trays were moved to 4°C for cryoprotection. Saturated LiSO4 was mixed with 80, 60, 40, 20, and 0% volume to volume ratio well solution, and individual crystals in cryo loops were passed through drops of these solutions in the same order and then plunge frozen in liquid N2. Anomalous wavelength anomalous diffraction at 3.28 Å revealed the complete polypeptide chain path and all disulfide bridges by about 50%. MolProbity was used for model validation.25

Heavy atom location, phasing, model building, and refinement

Data from 2 isomorphous crystals were indexed and integrated separately with Data from 2 isomorphous crystals were indexed and integrated separately with XDS and scaled together with XSCALE.17 Friedel pairs were kept unmerged. PHENIX-AUTOSOL used HySS for heavy atom location, PHASER to calculate phases using all refections to 3.28 Å, and allowed us to refine the structure to a low R free of 24.8% (Table 1). The entire CTCK domain and an N-acetyl glucosamine residue attached to residue Asn-2790 are resolved (Figure 1A). The asymmetric unit contains only one monomer. When viewing the structure with molecular graphics, the crystallographic dimer must be generated by adding a symmetry-related monomer.

The CTCK monomer adopts a highly elongated β-strand structure with dimensions of 6 x 3 x 2 nm (Figure 1A). The monomer is formed by 4 ribbons that alternate in direction as they go from one end of the monomer to the other. Ribbons (Figure 1B, right) are defined as going in a single direction from one end of a monomer to another and may contain one or more β-strands. Ribbon 1 contains an

### Table 1. Crystal diffraction data and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Wavelength (Å)</th>
<th>1.28237</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>47-3.28 (3.37-3.28)*</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P4₃2</td>
<td></td>
</tr>
<tr>
<td>Unit cell</td>
<td>a=b=γ=135.34 Å, α=β=γ=90°</td>
<td></td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>Total unique reflections</td>
<td>12294</td>
<td></td>
</tr>
<tr>
<td>Redundancy</td>
<td>13.7 (10.7)</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.99 (99.85)</td>
<td></td>
</tr>
<tr>
<td>Rmerge</td>
<td>10.87 (0.43)</td>
<td></td>
</tr>
<tr>
<td>Rfree</td>
<td>0.228 (6.097)</td>
<td></td>
</tr>
<tr>
<td>CC1/2 (%)</td>
<td>99.8 (12.0)</td>
<td></td>
</tr>
</tbody>
</table>

**Phasing and refinement**

*Phasing figure of merit | 0.245 |
| Rmerge / Rfree | 0.224 / 0.248 |
| Monomer / asymmetric unit | 1 |
| Nonhydrogen atoms | 721/14/0.5/25/6 |
| RMSD bonds (Å) | 0.003 |
| RMSD angles (°) | 0.46 |
| Ramachandran plot | 94.6/4.5/0.0 |
| Geometry and clash percentiles | 100% / 100% |
| Protein data bank ID | 4NT5 |

**Electron microscopy (EM) image processing**

Class averages of A3-CK were recalculated with centering on the CTCK region using a previously described auto-centering script (supplemental Figure 4).

### Results

#### Crystal structure of CTCK monomer

The VWF CTCK domain structure was solved using zinc single-wavelength anomalous diffraction at 3.28 Å. The experimental electron density, which was unusually good for this resolution, revealed the complete polypeptide chain path and all disulfide bridges and allowed us to refine the structure to a low R free of 24.8% (Table 1). The entire CTCK domain and an N-acetyl glucosamine residue attached to residue Asn-2790 are resolved (Figure 1A). The asymmetric unit contains only one monomer. When viewing the structure with molecular graphics, the crystallographic dimer must be generated by adding a symmetry-related monomer.

The CTCK monomer adopts a highly elongated β-strand structure with dimensions of 6 x 3 x 2 nm (Figure 1A). The monomer is formed by 4 ribbons that alternate in direction as they go from one end of the monomer to the other. Ribbons (Figure 1B, right) are defined as going in a single direction from one end of a monomer to another and may contain one or more β-strands. Ribbon 1 contains an...
N-terminal segment and the β1 and β2 strands, ribbon 2 contains the β3 and β4 strands, ribbon 3 is comprised only of the extremely long β5 strand, and ribbon 4 contains the β6 and β7 strands (Figure 1B). The 4 ribbons hydrogen bond into 2 β-ribbons. The 2 β-ribbons are also knit together by backbone hydrogen bonds of β3 and β4 with opposite ends of the β5 strand (Figure 1B). β-Ribbons contain only 2 antiparallel β-strands, and this building unit enables considerable twisting over the long lengths of the β4, β5, and β6 strands (Figure 1A).

### Intra-chain disulfides

The experimental electron density defines the connectivity of 4 intra-chain disulfides (Figure 1E-F) and completely agrees with chemical determination. In the descriptions below and in the figures, we refer to the cysteines by their order in the CTCK domain, e.g., C1 to C11, or by their pre-pro VWF sequence number, for example, Cys-2724 to Cys-2811 (Figure 1D).

The CK ties together the 4 ribbons at the middle of the long axis of the monomer (Figure 1A). Cys-2750 (C3) and Cys-2754 (C4) in ribbon 2 disulfide bond to Cys-2804 (C9) and Cys-2806 (C10), respectively, in ribbon 4 (Figure 1A-D). The C3-C9 and C4-C10 disulfides thus link β4 and β6, which run parallel to one another and do not hydrogen bond to one another. Linkage of C3 and C4, 5 residues apart in β4, to C9 and C10, 3 residues apart in β6, forms an 8-residue ring (Figure 1D). A third disulfide penetrates this ring, links Cys-2724 (C1) in ribbon 1 to Cys-2774 (C7) in ribbon 3 (Figure 1E),
and thus secures the CK. The fourth intra-chain Cys-2739 to Cys-2788 (C2-C8) disulfide further links the 2 β-ribbons at their tips, distal from the dimer interface (see below).

**Relation to CK cytokine monomers**

Searches for structural homologs of the CTCK monomer using DALI confirmed the uniqueness of the CTCK domain and its relationship to CK cytokines (supplemental Figure 5). CK cytokines usually contain a CK and 2 β-ribbons. However, the paths taken by each β-ribbon, the length of the β-ribbons, hydrogen bonding between the 2 β-ribbons, and presence of α-helices are highly variable. The cytokine members of the family all dimerize, whereas others such as sclerostin are monomers. Some contain N-terminal prodomains such as TGFβ; however, the functional forms of all previously described structural relatives function with only a single domain in their monomers; ie, none are mosaic proteins. The CTCK domain has not been previously structurally characterized, in agreement with the absence of any of the mosaic proteins described by Bork in structural homology searches.

**Dimerization of CTCK**

Each CTCK monomer interacts across three-quarters of its long axis to form the CTCK dimer (Figure 2). The twisted β4/β5 ribbon extends deeply into the other monomer in the dimer and forms extensive main chain hydrogen bonds (Figure 1B) and hydrophobic interactions (Figure 3A). The buried interface is quite large, covers 32% of the surface (1520 Å²) of each monomer, and includes inter-chain disulfides, backbone β-sheet hydrogen bonds, and sidechain complementarity.

Backbone hydrogen bonds across the symmetry axis extend the β-ribbons. Long β4, β5, and β6 strands that run through the dimer interface link into a super β-sheet, where β4 mates with β4 in the other monomer across the dyad axis (Figure 1B). Residues contributed by the other monomer are shown in gray in Figure 1B. β strand residues 2750-2758 form antiparallel hydrogen bonds to residues 2750-2758 in β9 (Figures 1B and 3A). Furthermore, a short β-ribbon is formed by β7, which interacts only with the symmetry-related β9 strand (Figures 1B and 2A).

Despite the long and narrow shape of the dimer interface, it contains significant burial of hydrophobic residues. Where the edges of the β4 strands meet to hydrogen bond across the dimer interface, Met-2759 forms hydrophobic interactions to Tyr-2749 across the interface (Figure 3A). Continuing along, Tyr-2760 and Ile-2762 tuck the end of the β4-β5 loop securely into a hydrophobic cavity on the other monomer, and Val-2767 in the β5 strand tucks into the same cavity (Figure 3A).

The structure shows that the 3 previously unassigned cysteines, Cys-2771, Cys-2773, and Cys-2811, all form inter-chain disulfides (Figure 2A-B, E). Cys-2771 and 2773 each reside in β-strand 5 and form reciprocal C2771-C2773 and C2773-C2771 disulfides. These β5-β5 inter-chain disulfides complement the extensive intra-chain hydrogen bonds formed by β5 at the interface between the 2 β-ribbons in each monomer (Figure 1B). Cys-2811 is at the center of
the short, 3-residue β7 strand, and disulfide bonds to its Cys-2811 counterpart in the β7’ strand to further secure the C-terminus of the CTCK domain.

In a highly force-resistant structure, the 3 inter-chain disulfides are sandwiched between the 3 CK disulfides in each monomer (Figure 2B). Thus, 9 pairs of disulfides all lie within an ellipsoid only 2.5 nm long and 1 to 1.5 nm in diameter. Of these 18 cysteines, all but 2 lie within β-strands. Furthermore, Cys-2771 and Cys-2773 that form the reciprocal inter-chain β5-β5’ disulfides lie adjacent to Cys-2774 in β5, which forms the loop-penetrating CK disulfide. The density of disulfide and β-sheet backbone crosslinks in this region is most remarkable (shown in stereo in Figure 3C). Adding to this, Ser-2756 in the β4-strand forms an unusually strong 2.4 Å sidechain hydrogen bond to the backbone between Cys-2773 and Cys-2774 (Figure 3C). Adding to this, Ser-2756 in the β4-strand forms an unusually strong 2.4 Å sidechain hydrogen bond to the backbone between Cys-2773 and Cys-2774 (Figure 3C). Notably, this highly crosslinked region lies in the center of the CTCK dimer immediately below the monomers’ N termini (Figure 2A), which bear all of the elongational force transmitted through VWF concatemers as a consequence of hydrodynamic flow.

Domains in the C-terminal portion of VWF dimers zip up into dimeric bouquet-like structure at acidic pH values found in the Golgi and Weibel-Palade bodies (Figure 2C).13 The CTCK dimer forms the slightly curved base at the C-terminal end of these dimers (Figure 2C arrow). The curved shape of the CTCK dimer crystal structure in the orientation shown in Figure 2A, and its long dimension of 8 nm, agree perfectly with its shape and dimension in negatively stained EM class averages (Figure 2D). Note that in the orientation of Figure 2A the N terminus of each CTCK monomer points vertically in an optimal orientation to connect to the C terminus of the VWC6 module in dimeric bouquets.

**VWD mutations**

VWD mutations in the CTCK domain can cause quantitative decrease in multimers (type 1), complete deficiency (type 3), or selective loss of longer multimers (type 2).27 In agreement with the importance of disulfide bonds to CTCK domain structure, 8 of 12 documented mutations27 occur in cysteines, and these cysteines contribute to all classes of disulfides, ie, the C2-C8 and CK intra-chain disulfides and the inter-chain disulfides (Figure 3B).

Interestingly, C2771S, C2771Y, C2773R, and C2773S are all type 2 mutations. These cysteines form the reciprocal Cys-2771-Cys-2773 and Cys2771’-Cys-2773 inter-chain disulfides. Mutation of either of these cysteines has the interesting property of disrupting not 1 but 2 of the 3 inter-chain disulfides, explaining the type 2 phenotype of selective loss of longer VWF concatemers. In contrast, mutations of cysteines forming intra-chain disulfides, C2739Y, C2754W, C2804R, and C2804Y, cause complete deficiency of VWF (type 3).

Mutations of 4 non-cysteine residues, including 3 prolines, cause type 1 and 2 VWD. The A2801D mutation introduces an aspartic acid sidechain into the center of the hydrophobic core of the dimer interface that cradles the β4-β5 ribbon and its loop (Figure 3A-B). This mutation should not disrupt monomer structure and in agreement with this, the P2776L mutation affects a Pro that is on the edge of the same hydrophobic cavity (Figure 3B) and helps...
mediate a twist in the β5-β6 ribbon, where its hydrogen bond pattern is disrupted (Figure 1B). P2776L causes type 1 VWD. P2772A is in the linker between the VWC6 and CTCK domains (Figure 3B). Only 4 residues, in a Glu-Glu-Pro-Glu sequence, intervene between the last Cys of VWC6 and first Cys of CTCK. P2772A is likely to affect the interaction between these domains and also causes type 1 VWD. Pro 2781, like Pro 2776, is in a region of the β5-β6 ribbon where hydrogen bonding is temporarily interrupted (Figure 1B). The P2781S mutation may disrupt this ribbon and causes type 2 VWD.

**Discussion**

Our crystal structure reveals the fold and mechanism of dimerization for the CTCK domain family. The structure provides insights into how the CTCK domain in VWF mediates formation of dimers in the endoplasmic reticulum. The dimerization interface is heavily reinforced with hydrogen bonds and 3 disulfide bonds and is sandwiched between the similarly heavily reinforced CK regions in each monomer. The inter-chain disulfide bond linkage is in agreement with 1 of 2 models previously proposed based on exclusion of intra-chain linked cysteines, structural models, mutation, and symmetry arguments. The structure explains VWD mutation phenotypes. Mutations of intra-chain linked cysteines cause complete VWF deficiency and mutations of inter-chain linked cysteines cause selective deficiency of large multimers. These findings suggest that intra-chain disulfide linkage in monomers precedes and is required for subsequent inter-chain linkage in dimers.

We propose that the highly crosslinked dimer interface in CTCK is a specialization for force resistance. Most investigators are familiar with the concept that DNA is sufficiently long for its covalent bonds to be broken in the shear flow of a pipette. Similarly, VWF is sufficiently long to have significant hydrodynamic force exerted on it. Even free in flow, the hydrodynamic force on VWF is sufficient to unfold the A2 domain. When bound to the vessel wall and platelets, the force on VWF would be much greater. An elementary principle of proteins, first demonstrated by Anfinsen, is that protein folding favors specific disulfide formation. Thus, disulfides that are buried in a protein core remain intact as disulfides even in the presence of reducing agents. However, reduction occurs when denaturants are added or tensile force is applied. In plasma, the concentrations of glutathione/oxidized glutathione and cysteine/cystine are 0.14/2.8 and 10/40 μM, respectively. When force is applied across a disulfide bond, the kinetics of its reduction greatly increase. Reduction of the inter-chain disulfides in the CTCK domain (or D3 domain) in a force-elongated VWF concatemer would have disastrous consequences, because the 2 half-molecules would separate in flow and have little chance of finding one another and reannealing after force subsided. In contrast, reduction of a long-range disulfide in an internal domain of VWF such as A1 could easily be subsequently reversed by refolding and oxidation.

We thus reemphasize the highly reinforced structure of the CTCK dimer. Ten inter-chain backbone hydrogen bonds link the β4 strands in each monomer to form a β6-β5-β4-β4'-β5'-β6' supersheet, and 4 hydrogen bonds further link the β7 and β7' strands. These interactions are enhanced by burial of each β4-β5 ribbon and its loop in a hydrophobic pocket in the other monomer. All but one of 6 CK cysteines, and all 3 cysteines that form inter-chain disulfides, are present in β-strands. This is important, because β-structures are rigid compared with α-helices and loops and thus better suited for force resistance. If a protein deforms in response to force, it is more readily elongated and unfolded. A term corresponding to compliance appears in an exponent along with force and determines how much force exponentiates the rate of unfolding. Finally, the β-structure in the inter-chain disulfide region is continuous with that in the CK region of each monomer, which immediately flanks each side of, and further reinforces, the inter-chain disulfide region. Having each looked at a large number of extracellular protein modules, we have never seen a domain of a soluble extracellular or membrane protein, let alone a dimer interface, so bristling with disulfide bonds and backbone hydrogen bonds (Figures 1B and 3C). The only other such heavily reinforced proteins that come to mind are incorporated into

---

**Figure 4. Diverse dimerization mechanisms among VWF CTCK and its relatives.** Dimer structures are superimposed on the CTCK monomer in cyan and then aligned with it horizontally and vertically on the page. Disulfide bonds are shown in stick and colored as in Figure 1. (A) VWF CTCK. (B) Human chorionic gonadotropin (PDB code 1QFW). (C) Neurotrophin-4 (PDB 1HCF). (D) Noggin (PDB 1M4U). (E) TGFβ2 (PDB 1TFG). (F) PDGF (PDB 1PDG).
the extracellular matrix, such as the noncollagenous NC1 domain of collagen IV, which mediates collagen crosslinking into mechanically tough sheets.33

The uniqueness of the CTCK dimerization interface is further emphasized by comparison with CK cytokines (Figure 4). CK dimerization mechanisms are diverse with 5 different types of orientations (Figure 4B-F). Chorionic gonadotropin and follicle stimulating hormone dimerize with their monomers antiparallel, the same spacing between the CKs as in CTCK, and are the only other group where monomer β-ribbons form a dimeric super-β-sheet as in VWF. However, this CK cytokine subfamily completely lacks inter-chain disulfides.

Members of the TGFβ superfamily dimerize in a similar orientation but have α-helices at the dimer interface (Figure 4E). There is one inter-chain disulfide, mediated by the equivalent of VWF Cys-2773. However, in the TGFβ family this cysteine dimerizes with the identical cysteine in the counterpart monomer (Figure 5) rather than to a cysteine 2 residues away in the same β-strand as in VWF. Therefore, the dimer interface slides in the TGFβ family, so the CKs in the 2 monomers are 2 β-ladder positions closer than in CTCK (Figure 4E).

In noggin, a long helix-loop-helix N-terminal addition to the CK domain contributes most of the dimerization interface (Figure 4D). A 2-residue insertion between knot cysteines C3 and C4 (Figure 5) disrupts the hydrophobicity of the inter-monomer interface. Nonetheless, noggin achieves an inter-monomer orientation different from CTCK and a cysteine in a similar sequence position to C11 in CTCK (Figure 5) mediates the inter-chain disulfide (Figure 4D).

Figure 5. Sequence alignment of VWF CTCK to CTCK domains from other proteins and structural alignment to related proteins. Sequence alignment with MAFFT was with the G-INS-i strategy. Structural alignment was with SSM; the alignment was then condensed, closing gaps in loop regions and preserving alignment in β-strands (overlined) and in cysteines. Sequence accessions are murine mucin5AC, GI:114431224; zebrafish otogelin, GI:326669509; Drosophila hemolectin, GI:24663920; human norrin, GI:4557789; Drosophila Silt, GI:17136842; human connective tissue growth factor, GI:49456477; human TGFβ1, GI:48429157; human noggin, GI:1117817; and human chionic gonadotropin β-subunit, GI:132566338.

How do other CT family members compare with CTCK in VWF?

The VWF CTCK subfamily includes hemolectin, norrin, mucins, and otogelin (Figure 5). These are predicted to dimerize similarly to VWF CTCK. All contain sequence signatures for the long β-strands present in CTCK, all contain C5 and C6 in β that form the reciprocal inter-chain disulfides, all have C2 and C8 that form the intrachain disulfide unique to CTCK domains, and most contain C11 that forms the C11-C11' inter-chain disulfide (Figure 5). Differences among the CTCK family are easily interpretable from our structure-sequence alignment.

The hemolectin/hemocytin proteins in insects are secreted by hemocytes into hemolymph.34 They contain all of the modules found in VWF, together with discoidin domains that recognize carbohydrates, and function in both clotting and innate immune defense. The insertion in hemolectin (Figure 5) will extend the β2 and β3 strands and/or their loop, which lie distal from the dimer interface.

Mucins are also evolutionarily ancient and function in development in the ear (otogelin) as well as in mucous barriers. Many gel-forming mucins contain the CTCK domain. In otogelin, lack of the β7 strand and its inter-chain C11 cysteine (Figure 5) is easily accommodated structurally. Like VWF, mucins with CTCK domains contain N-terminal D domains and are assembled into multimers, although these may be net-like hexagonal arrays rather than linear.11,35 Mucins may also bear high forces, both as a consequence of their high viscosity.
mutation of the cysteine equivalent to C6, which disulfide bonds across the dimer interface to C5 in VWF, results in dimers. If the C6-C5’ and C11-C11’ interchain disulfides were among different rather than identical monomers in norrin, ie, C6-C5’ and C11-C11’), then multimers could form. Norrin signals in multiple developmental pathways by binding frizzled4. The binding site has been mapped by alanine scanning mutations and modeled using a BMP2 dimer. Our pathways by binding frizzled4. The binding site has been mapped by alanine scanning mutations and modeled using a BMP2 dimer. Our VWF CTCK dimer structure provides an improved modeling template. Mutations in norrin define a contiguous frizzled4 binding site in ribbon 1, the β3-β4 loop, the C-terminus of β5, and the middle of β6 (supplemental Figure 6).

The Slit and CCN families regulate axon guidance and diverse adhesive and signaling activities in the extracellular matrix, respectively. Exemplified by Slit1 and connective tissue growth factor in Figure 5, they are predicted to have structures intermediate between the VWF CTCK domain and TGFβ. Slit and CCN family members lack cysteine C5, and thus their dimerization interface is predicted, like that in TGFβ, to slide 2 residues relative to CTCK. This is consistent with shortening of β4 and β5 in Slit and CCN, because these tuck into the dimer interface, and sliding leaves less room for this interface. Slit, connective tissue growth factor, and their relatives share the C2, C8, and C11 cysteines with VWF CTCK and not with CK cytokine family members and therefore are predicted to share the C2-C8 intrachain and C11-C11’ interchain disulfides with VWF. Thus, Slit and CCN members appear to have a dimerization interface that is a hybrid of the interface found in the VWF CTCK domain and TGFβ superfamily. This may require future adjustment of CTCK domain nomenclature to split CTCK domains into 2 subfamilies.

In conclusion, the CTCK domain structure provides specific insights into an important dimerization and force-bearing module in VWF and a more general overview of a dimerization module in proteins with surprisingly diverse functions.

Authorship

Contribution: Y.-F.Z. performed research, analyzed data, designed research, and wrote the manuscript; and T.A.S. analyzed data, designed research, and wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Timothy A. Springer, Program in Cellular and Molecular Medicine, Children’s Hospital Boston and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115; e-mail: timothy.springer@childrens.harvard.edu.

References

34. Lesch C, Goto A, Lindgren M, Bidda G, Dushay MS, Theopold U. A role for Hemoeclin in coagulation and


Highly reinforced structure of a C-terminal dimerization domain in von Willebrand factor

Yan-Feng Zhou and Timothy A. Springer